Halloween Candy

Andres Rivero

10/29/2021

```
candy_file <- "candy-data.txt"
candy = read.csv(candy_file, row.names=1)
head(candy)
```

```
##                chocolate fruity caramel peanutyalmondy nougat crispedricewafer
## 100 Grand              1      0       1              0      0                1
## 3 Musketeers           1      0       0              0      1                0
## One dime               0      0       0              0      0                0
## One quarter            0      0       0              0      0                0
## Air Heads              0      1       0              0      0                0
## Almond Joy             1      0       0              1      0                0
##                hard bar pluribus sugarpercent pricepercent winpercent
## 100 Grand         0   1        0        0.732        0.860   66.97173
## 3 Musketeers      0   1        0        0.604        0.511   67.60294
## One dime          0   0        0        0.011        0.116   32.26109
## One quarter       0   0        0        0.011        0.511   46.11650
## Air Heads         0   0        0        0.906        0.511   52.34146
## Almond Joy        0   1        0        0.465        0.767   50.34755
```

```
dim(candy)
```

```
## [1] 85 12
```

Q1 85 different types of candy

```
sum(candy$fruity)
```

```
## [1] 38
```

Q2 38 of them

```
candy["Kit Kat", ]$winpercent
```

```
## [1] 76.7686
```

Q3 Kit Kat, 76%

Q4 76%

Q5

```
candy["Tootsie Roll Snack Bars", ]$winpercent
```

```
## [1] 49.6535
```
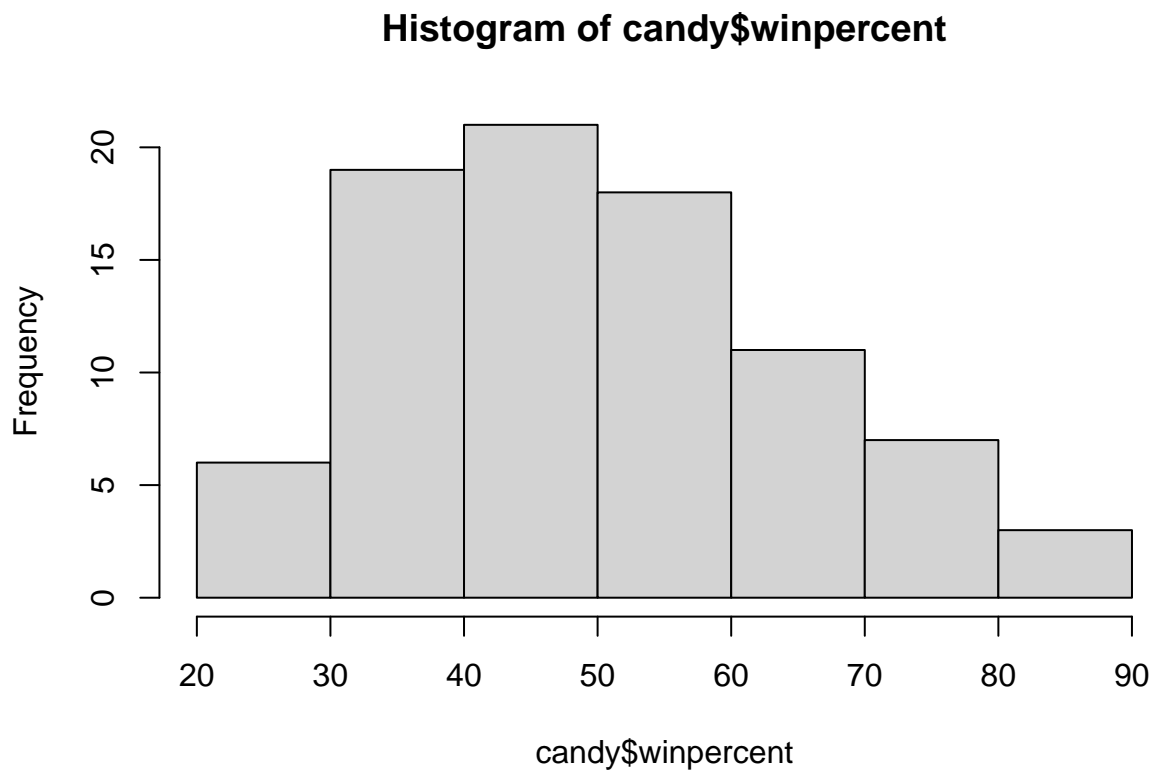
49%

```
#install.packages("skimr")
#library("skimr")
```

```
#skim(candy)
```

Q6 Yes, the histogram column for example

Q7 0 means it doesn't have it or is not considered to be chocolate candy, and 1 means it is considered to have chocolate.

Q8

```
hist(candy$winpercent)
```


Histogram of candy$winpercent

Q9 No it isn't.

Q10 Below

Q11

```
chocolate <- candy[as.logical(candy$chocolate),]$winpercent
mean(chocolate)
```

```
## [1] 60.92153
```

```
fruity <- candy[as.logical(candy$fruity),]$winpercent
mean(fruity)
```

```
## [1] 44.11974
```

Chocolate is higher

Q12

```
t.test(chocolate, fruity)
```

```
##
##  Welch Two Sample t-test
##
## data:  chocolate and fruity
## t = 6.2582, df = 68.882, p-value = 2.871e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  11.44563 22.15795
## sample estimates:
## mean of x mean of y
##  60.92153  44.11974
```

Q13

```
head(candy[order(candy$winpercent),], n=5)
```

```
##                   chocolate fruity caramel peanutyalmondy nougat
## Nik L Nip                 0      1       0              0      0
## Boston Baked Beans        0      0       0              1      0
## Chiclets                  0      1       0              0      0
## Super Bubble              0      1       0              0      0
## Jawbusters                0      1       0              0      0
##                   crispedricewafer hard bar pluribus sugarpercent pricepercent
## Nik L Nip                        0    0   0        1        0.197        0.976
## Boston Baked Beans               0    0   0        1        0.313        0.511
## Chiclets                         0    0   0        1        0.046        0.325
## Super Bubble                     0    0   0        0        0.162        0.116
## Jawbusters                       0    1   0        1        0.093        0.511
##                   winpercent
## Nik L Nip            22.44534
## Boston Baked Beans   23.41782
## Chiclets             24.52499
## Super Bubble         27.30386
## Jawbusters           28.12744
```

Q14

```
tail(candy[order(candy$winpercent),], n=5)
```

```
##                          chocolate fruity caramel peanutyalmondy nougat
## Snickers                         1      0       1              1      1
## Kit Kat                          1      0       0              0      0
## Twix                             1      0       1              0      0
## ReeseÃ•s Miniatures              1      0       0              1      0
## ReeseÃ•s Peanut Butter cup       1      0       0              1      0
##                          crispedricewafer hard bar pluribus sugarpercent
## Snickers                                0    0   1        0        0.546
## Kit Kat                                 1    0   1        0        0.313
## Twix                                    1    0   1        0        0.546
## ReeseÃ•s Miniatures                     0    0   0        0        0.034
## ReeseÃ•s Peanut Butter cup              0    0   0        0        0.720
##                          pricepercent winpercent
## Snickers                        0.651   76.67378
## Kit Kat                         0.511   76.76860
## Twix                            0.906   81.64291
## ReeseÃ•s Miniatures             0.279   81.86626
## ReeseÃ•s Peanut Butter cup      0.651   84.18029
```

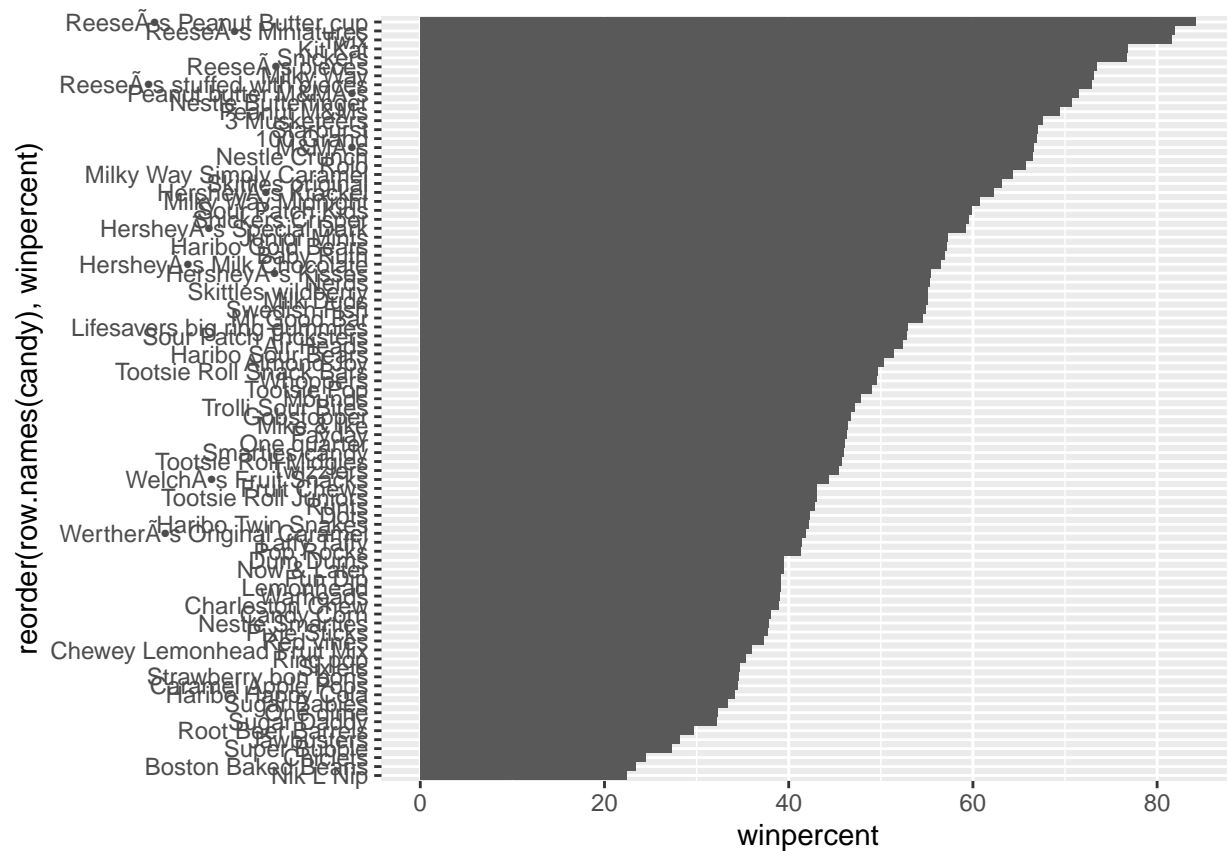**Q15**

```
library(ggplot2)

ggplot(candy) +
  aes(winpercent, rownames(candy)) +
  geom_col()
```
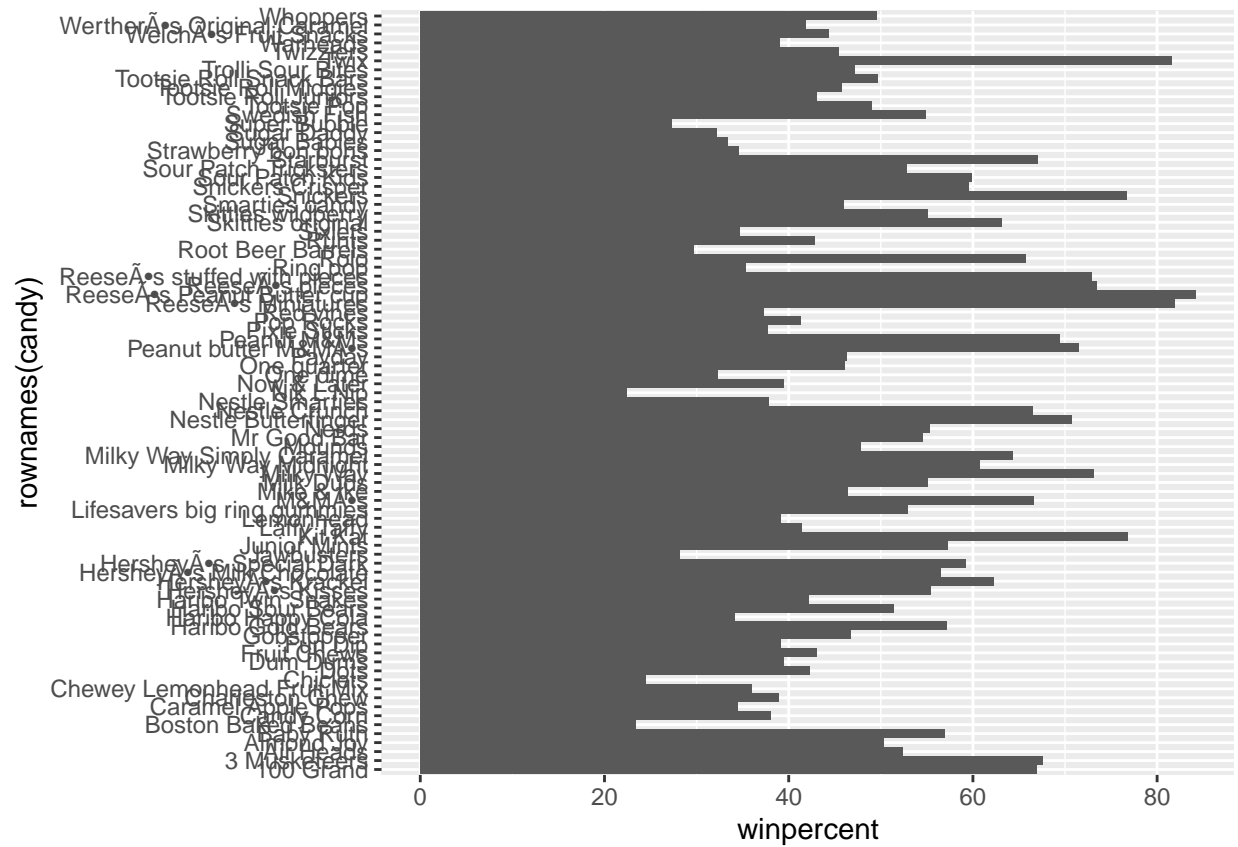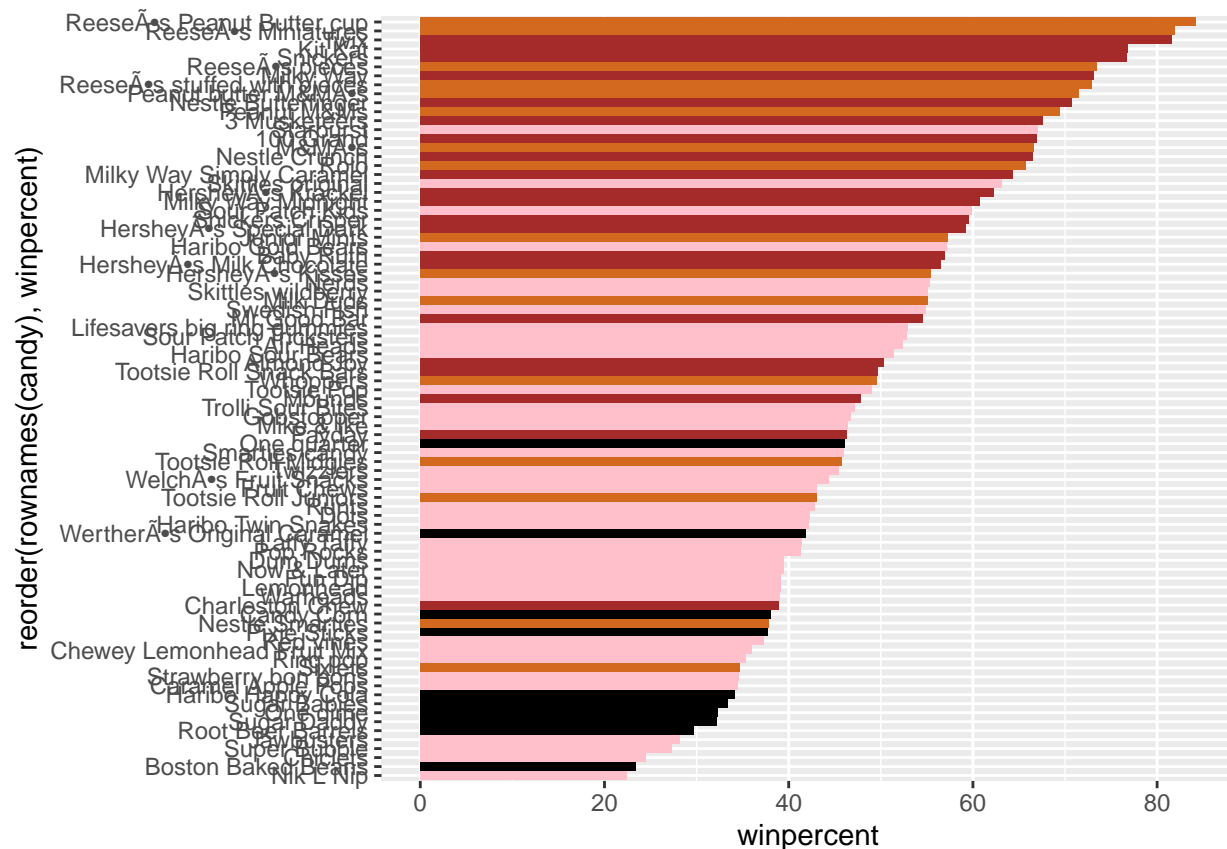
Q6

```
library(ggplot2)

ggplot(candy) +
  aes(winpercent, reorder(row.names(candy), winpercent)) +
  geom_col()
```

```
my_cols=rep("black", nrow(candy))
my_cols[as.logical(candy$chocolate)] = "chocolate"
my_cols[as.logical(candy$bar)] = "brown"
my_cols[as.logical(candy$fruity)] = "pink"
```

```
ggplot(candy) +
  aes(winpercent, reorder(rownames(candy),winpercent)) +
  geom_col(fill=my_cols)
```

Q17 Charleston Chew

Q18 Starburst

Q19

#install.packages("ggrepel") #library("ggrepel")

#ggplot(candy) + #aes(winpercent, pricepercent, label=rownames(candy)) + #geom_point(col=my_cols) + #geom_text_repel(col=my_cols, size=3.3, max.overlaps = 5)

Reeses Miniatures

Q20

```
ord <- order(candy$pricepercent, decreasing = TRUE)
head( candy[ord,c(11,12)], n=5 )
```

```
##                        pricepercent winpercent
## Nik L Nip                     0.976   22.44534
## Nestle Smarties               0.976   37.88719
## Ring pop                      0.965   35.29076
## HersheyÃ•s Krackel            0.918   62.28448
## HersheyÃ•s Milk Chocolate     0.918   56.49050
```

```
#install.packages("corrplot")
#library(corrplot)
#cij <- cor(candy)
#corrplot(cij)
```

Q22 Chocolate and Fruity

Q23 Chocolate and Bars

Principle Component Analysis

```
pca <- prcomp(candy, scale=TRUE)
summary(pca)
```

```
## Importance of components:
##                           PC1    PC2    PC3     PC4    PC5     PC6     PC7
## Standard deviation     2.0788 1.1378 1.1092 1.07533 0.9518 0.81923 0.81530
## Proportion of Variance 0.3601 0.1079 0.1025 0.09636 0.0755 0.05593 0.05539
## Cumulative Proportion  0.3601 0.4680 0.5705 0.66688 0.7424 0.79830 0.85369
##                           PC8     PC9    PC10    PC11    PC12
## Standard deviation     0.74530 0.67824 0.62349 0.43974 0.39760
## Proportion of Variance 0.04629 0.03833 0.03239 0.01611 0.01317
## Cumulative Proportion  0.89998 0.93832 0.97071 0.98683 1.00000
```